

On the Chernoff bound for efficiency of quantum hypothesis testing

Vladislav Kargin*

Abstract

The paper estimates the Chernoff rate for the efficiency of quantum hypothesis testing. For both joint and separable measurements, approximate bounds for the rate are given if both states are mixed and exact expressions are derived if at least one of the states is pure. The efficiency of tests with separable measurements is found to be close to the efficiency of tests with joint measurements. The results are illustrated by a test of quantum entanglement.

1 Introduction

Mark Kac once called Probability Theory measure theory with a “soul” provided by Physics, games of chance, Economics or Geometry.¹ In a sense then, Quantum Statistics can be called probability theory with a “subconscious”. The probability distributions, so important for classical statistics, are no longer the deepest layer of foundations but only an outward manifestation of geometry in the Hilbert spaces of quantum states. This foundational change begs for a new look at the classical statistics results, and this paper contributes by reconsidering the Chernoff-Hoeffding results about hypothesis testing.

Why quantum statistics? Today, quantum states can be manufactured. For example, in one method (Cirac and Zoller (1995)) ions are placed in a trap created by electrostatic potential and radio-frequency oscillations. The ions then are cooled by laser emission, and arranged on a line in the trap. After that, each individual ion can be accessed by laser pulses and their joint quantum state can be altered according to the researcher’s wishes. This ability to built and manipulate quantum systems is changing our thinking about computation and information transmission. Suddenly, certain classic problems – the factorization of large integers, the search in an unstructured database, secure communication – are not as difficult as they always were.

This conceptual change also affects statistics.

For example, how can a quantum state manufacturer check if states have been generated faithfully? We can anticipate the statistician’s answer: Select a

*Cornerstone Research, 599 Lexington Avenue, New York, NY 10022, USA; slava@bu.edu

¹In a preface to a book about geometric probability by Santaló (1976).

sample of the states and perform a statistical test. But now, besides designing the test, the statistician must play an additional role, the role of advisor on how to perform measurements of a sample of quantum states. Since in quantum mechanics both the measurement and the state determine the probability distribution of outcomes, the choice of measurement affects the properties of the statistical test.

Not all measurements are readily available. Sometimes it is possible to measure sample states jointly, as one large quantum state, and sometimes the states can only be measured separately and simultaneously. Yet another possibility is that the states must be measured separately and sequentially. Finally, sometimes the sample states can only be measured partially, for example, when each state represents several remote particles that cannot be measured jointly. Clearly, the efficiency of the optimal test will depend on which measurements are available. In this paper we will concentrate on joint and separable independent measurements.

For a single state the problem of quantum hypothesis testing was solved by Holevo (1976) and Helstrom (1976). In this paper, I consider a different situation: when the researcher has access to several copies of the same state but may not be able to measure them jointly.

The problem of testing using a sample of states was also considered in Helstrom (1976) (see also recent results by Ogawa and Nagaoka (2000) and Parthasarathy (2001)). These authors considered only joint measurements and only the situation when one of the errors may go to zero arbitrarily slowly. In contrast, I consider a Bayesian version of the problem, in which the researcher aims to minimize a weighted average of both errors, and I consider both joint and separable measurements.

When joint measurements are available, the problem of testing using a sample can be solved by applying the Holevo-Helstrom result to the case of tensor powers of primary states. In this case, my main results provide useful bounds on both the expected error when the sample is finite and the rate of decline in error as the number of sample states grows. The bounds are given in terms of fidelity distance between quantum hypothesis.

In addition I derive explicit expression for the rate of error decline in cases if either one of the hypothesis specifies pure state, or the states specified by the hypotheses commute.

For the separable measurements, I mainly concentrate on the asymptotic case. If at least one of the hypotheses is pure, then the optimal separable measurement has the same asymptotic error rate as the joint measurement. If both hypotheses are mixed, then there is a measurement whose performance is close to the performance of the joint optimal measurement. Together these results imply that the loss in efficiency associated with restriction on available measurements is not large.

This paper contributes only to the theory of quantum hypothesis testing. I do not touch on another rapidly growing area of research: quantum state estimation. For recent progress in this area see the review article by Barndorff-Nielsen, Gill, and Jupp.

The rest of the paper is organized as follows. Section 2 gives some basic in-

formation about quantum states and measurements and formulates the problem of quantum hypothesis testing. Section 3 gives a short summary of the Chernoff-Hoeffding results about hypothesis testing. Sections 4 and 5 discuss joint and separable measurements, respectively. Section 6 presents an illustration. And Section 7 concludes.

2 Quantum Hypothesis Testing

States of quantum-mechanical objects – electrons, photons, atoms, molecules, etc. – are described by density matrices. A density matrix is a self-adjoint, non-negative operator of a complex Hilbert space with a trace of 1. In this paper we will be concerned only with finite-dimensional Hilbert spaces, so the operator is indeed represented by a finite Hermitian matrix. A particular case is matrices of rank one. They are projectors on one-dimensional subspaces and called pure states.

States are not directly observable: they can be measured but the outcome of a measurement is a random variable. More precisely, measurements are sets of non-negative operators which are required to add up to the identity operator. Each operator corresponds to a particular outcome of the measurement, and the probability of outcome i if the state is ρ and the measurement is $\{M_i\}$ is $\text{tr}\{M_i\rho\}$. An important subclass is formed by measurements in which the outcomes are orthogonal projectors: $M_i M_j = \delta_{ij} M_i$, where δ_{ij} is the Dirac delta-function.

We consider the following problem: a researcher is given a sample of N identical quantum states, which are either ρ_0 or ρ_1 with the prior probability $1/2$. He aims to minimize the average probability of making an incorrect decision about the state by devising a system of measurements and a decision rule. Can we safely assume that all measurements are available to the researcher? No.

While in some situations the researcher can make a joint measurement of the state that represent the total sample, most often he can do only separate measurements of each state in the sample. If the measurements are done independently of each other, then we will call them separable independent measurements. If the measurements can be done sequentially and the researcher adjusts the current measurement according to the results obtained in the previous measurements, then they are separable adaptable measurements.

Sometimes, the researcher is even more restricted. This happens, for example, if a sample quantum state consists of two spatially remote parts and the researcher can only measure them separately. Mathematically it means that the operators of the measurement must be block-diagonal in a certain basis. This setup may raise interesting statistical issues about identification of the state properties.

3 Classical Chernoff-Hoeffding Bounds

This section reviews results by Chernoff (1952), Sanov (1957), and Hoeffding (1965) about asymptotic error rates in hypothesis testing. For details the reader can also consult the book by Cover and Thomas.

Consider two multinomial distributions, P and Q , from one of which a sample is drawn and provided to a researcher. The researcher's task is to guess the distribution. The sufficient statistic for this problem is the empirical distribution of the sample, X , and the decision rule is specified by two complementary sets, \mathcal{P} and \mathcal{Q} , of probability distributions on outcomes. If $X \in \mathcal{P}$, hypothesis P is accepted; otherwise, Q is accepted. It is assumed that $P \in \mathcal{P}$, and $Q \in \mathcal{Q}$.

If the true probability distribution is P , it is the Sanov theorem that asymptotically the probability of making an error and accepting Q is

$$\exp[-ND(\mathcal{Q}||P)], \quad (3.1)$$

up to a subexponential factor, where $D(\mathcal{Q}||P)$ is the Kullback-Leibler distance from P to \mathcal{Q} :

$$D(\mathcal{Q}||P) = \min_{S \in \mathcal{Q}} \sum_{i=1}^N s_i \ln \frac{p_i}{s_i}. \quad (3.2)$$

It follows that the average probability of making an error declines asymptotically with growth in N :

$$R \sim \exp[-N \min\{D(\mathcal{Q}||P), D(\mathcal{P}||Q)\}]. \quad (3.3)$$

The maximum of the decline rate over all possible \mathcal{P} and \mathcal{Q} is sometimes called the Chernoff information distance between distributions P and Q :

$$D_c(P, Q) = \max_{\mathcal{P}, \mathcal{Q}} \min\{D(\mathcal{Q}||P), D(\mathcal{P}||Q)\} \quad (3.4)$$

Hoeffding proved that the optimal sets \mathcal{P} and \mathcal{Q} can be determined from the maximum likelihood principle: a distribution S belongs to \mathcal{P} if and only if $D(S||P) \leq D(S||Q)$. Intuitively, in this case distribution S is more likely to be observed if the true distribution is P rather than if it is Q , so hypothesis P should be accepted.

For example, for multinomial distribution we have the following formula for the probability of error in the optimal test:

$$R = \frac{1}{2} \left\{ \sum_{\substack{p_1^{x_1} \dots p_n^{x_n} < q_1^{x_1} \dots q_n^{x_n} \\ x_1 + \dots + x_n = N}} p_1^{x_1} \dots p_n^{x_n} + \sum_{\substack{p_1^{x_1} \dots p_n^{x_n} > q_1^{x_1} \dots q_n^{x_n} \\ x_1 + \dots + x_n = N}} q_1^{x_1} \dots q_n^{x_n} \right\} \quad (3.5)$$

$$= \frac{1}{2} \left\{ 1 - \frac{1}{2} \sum_{x_1 + \dots + x_n = N} |p_1^{x_1} \dots p_n^{x_n} - q_1^{x_1} \dots q_n^{x_n}| \right\} \quad (3.6)$$

$$= \frac{1}{2} \left\{ 1 - \frac{1}{2} \|P_N - Q_N\|_1 \right\} \quad (3.7)$$

$$\sim c \exp[-ND_c(P, Q)], \quad (3.8)$$

where the sums are taken over possible results of sampling N times from a multinomial distribution with n outcomes: x_1 is the number of outcomes of type 1, x_2 is the number of outcomes of type 2, and so on; P_N and Q_N are distributions on the sample space induced by the distributions P and Q on outcomes, and $\|\cdot\|_1$ is the total variation norm.

It also turns out (see Cover and Thomas (1991) for derivation) that for the optimal choice of \mathcal{P} and \mathcal{Q} , the probability distributions $S \in \mathcal{Q}$ and $S' \in \mathcal{P}$ that minimize respectively $D(S||P)$ and $D(S'||Q)$ are the same and given by the following formula:

$$s_i = \frac{p_i^\lambda q_i^{1-\lambda}}{\sum_{j=1}^N p_j^\lambda q_j^{1-\lambda}}, \quad (3.9)$$

where λ is chosen in such a way that $D(S||P) = D(S||Q)$. Knowing expression (3.9) we can derive another expression for the asymptotic probability of error:

$$\frac{1}{N} \ln R = \min_{0 \leq \lambda \leq 1} \log \sum_{i=1}^N p_i^\lambda q_i^{1-\lambda}. \quad (3.10)$$

All these derivations presuppose that P and Q are fixed. In quantum statistics the researcher has the ability to vary P and Q by choosing the measurement. How does this change the classical results?

4 Joint Measurements

In this section we look at the joint measurements of a sample of quantum states. The minimal expected error obtained in this case is a lower bound on the error achievable when the set of measurements is restricted. In addition, the theory for joint measurements provides a fascinating counterpart to the classical theory of the Chernoff bounds.

4.1 Generalities

Joint measurement of all sample states is by definition a measurement of the tensor product of the sample states. Thus, in effect we have the problem of testing two alternative hypotheses about a single – although huge – quantum state, the problem that was solved by Holevo and Helstrom (see, for example, Holevo (2001)). In our situation we only need to determine what additional implications follow from the special structure of the state.

If the hypotheses about the quantum state are given by matrices ρ_0 and ρ_1 with prior probability of $1/2$, then according to the Holevo-Helstrom result, the optimal measurement is an orthogonal measurement with d outcomes, where d is the dimension of the Hilbert space and the outcomes are projectors on the eigenvectors of operator $\rho_0 - \rho_1$. The decision is made based on the following rule: If the measurement outcome corresponds to an eigenvector with a positive eigenvalue, then ρ_0 is chosen; otherwise, ρ_1 is chosen.

The minimal expected error probability that can be achieved after the optimal measurement is given by the following formula:

$$R = \frac{1}{2} \left(1 - \frac{1}{2} \|\rho_0 - \rho_1\|_1 \right), \quad (4.1)$$

where $\|\cdot\|_1$ denotes the sum of the absolute values of eigenvalues.

In our case the hypothetical states are tensor powers of the individual states, $\rho_0^{\otimes N}$ and $\rho_1^{\otimes N}$, where

$$\rho_i^{\otimes N} \equiv \underbrace{\rho_i \otimes \rho_i \otimes \dots \otimes \rho_i}_N. \quad (4.2)$$

The number of outcomes in the optimal joint measurement is d^N , so it can be enormous for large values of N . The error is

$$R = \frac{1}{2} \left(1 - \frac{1}{2} \|\rho_0^{\otimes N} - \rho_1^{\otimes N}\|_1 \right). \quad (4.3)$$

Note the similarity with classical expression (3.7).

What is the asymptotic rate of decline in error? Can we explicitly calculate the distribution of eigenvalues of $\rho_0^{\otimes N} - \rho_1^{\otimes N}$?

Initial moments of this distribution are indeed easy to calculate. Let us introduce a notation for the moments:

$$\mu_n =: \int_0^1 t^n dF(t) = \frac{1}{d^N} \text{tr} (\rho_0^{\otimes N} - \rho_1^{\otimes N})^n, \quad (4.4)$$

where $F(t)$ is the discrete probability distribution that puts equal probability weight on each eigenvalue. Then the following Proposition holds

Proposition 4.1

$$\mu_n = \frac{1}{d^N} \sum_{\{k_1, \dots, k_n\}} (-1)^{\sum k_i} (\text{tr} (\rho_{k_1} \dots \rho_{k_n}))^N, \quad (4.5)$$

where $\{k_1, \dots, k_n\}$ run over the set of all n -sequences of 0 and 1.

Proof: The proposition follows from the non-commutative binomial expansion of $(\rho_0^{\otimes N} - \rho_1^{\otimes N})^n$ and the fact that $\text{tr}(\rho_{k_1}^{\otimes N} \dots \rho_{k_n}^{\otimes N}) = (\text{tr}(\rho_{k_1} \dots \rho_{k_n}))^N$. QED

The advantage of this formula is that for a fixed n , the calculation is as easy for large as for small values of N . The difficulty is that the number of terms in this formula grows exponentially with n . Therefore the standard map from the set of moment sequences to the set of distributions is impractical. In the next sections we will pursue a different approach to estimation of $\|\rho_0^{\otimes N} - \rho_1^{\otimes N}\|_1$.

4.2 Special Cases

To get more insight about the behavior of $\|\rho_0^{\otimes N} - \rho_1^{\otimes N}\|_1$, it is useful to consider several special cases: (1) when both states are pure; and (2) when the density operators commute. In the first case let $\rho_0 = |\psi_0\rangle\langle\psi_0|$ and $\rho_1 = |\psi_1\rangle\langle\psi_1|$.² Then we have the following result:

Theorem 4.2 *If both states are pure, then the average error probability is*

$$R = \frac{1}{2} \left(1 - \sqrt{1 - |\langle\psi_0|\psi_1\rangle|^{2N}} \right). \quad (4.6)$$

Asymptotically,

$$R \sim \frac{1}{4} |\langle\psi_0|\psi_1\rangle|^{2N} \text{ as } N \rightarrow \infty. \quad (4.7)$$

Proof: Because of (4.1), we need only to prove that for pure states

$$\|\rho_0^{\otimes N} - \rho_1^{\otimes N}\|_1 = 2\sqrt{1 - |\langle\psi_0|\psi_1\rangle|^{2N}}. \quad (4.8)$$

We can write

$$\|\rho_0^{\otimes N} - \rho_1^{\otimes N}\|_1 = \left\| \left| \psi_0^{\otimes N} \right\rangle \left\langle \psi_0^{\otimes N} \right| - \left| \psi_1^{\otimes N} \right\rangle \left\langle \psi_1^{\otimes N} \right| \right\|_1. \quad (4.9)$$

Operator $\left| \psi_0^{\otimes N} \right\rangle \left\langle \psi_0^{\otimes N} \right| - \left| \psi_1^{\otimes N} \right\rangle \left\langle \psi_1^{\otimes N} \right|$ acts nontrivially only in a 2-dimensional space spanned by $\psi_0^{\otimes N}$ and $\psi_1^{\otimes N}$, and it is easy to compute the operator eigenvalues in this space. They are

$$\pm \sqrt{1 - \left| \left\langle \psi_0^{\otimes N} \right| \psi_1^{\otimes N} \right\rangle|^2} = \pm \sqrt{1 - |\langle\psi_0|\psi_1\rangle|^{2N}}. \quad (4.10)$$

From this and the fact that all other eigenvalues are zero, the first equality of the theorem follows. The asymptotic expression follows from the Taylor series for the square root.

QED.

Now consider the case of commuting ρ_0 and ρ_1 . Let the distributions of eigenvalues be P for ρ_0 and Q for ρ_1 .

Theorem 4.3 *If states commute, the average error probability is asymptotically*

$$R \sim c \exp[-ND_c(P, Q)] \quad (4.11)$$

In other words, the probability of error has exactly the same growth rate as in the classical case.

²For convenience, we use the Dirac ket-bra notation: the elements of the Hilbert space are denoted as $|\psi\rangle$, and the linear functionals on the Hilbert space are denoted as $\langle\psi|$. In particular, $|\psi_0\rangle\langle\psi_0|$ is the orthogonal projector on $|\psi_0\rangle$.

Proof: Since the density operators ρ_0 and ρ_1 commute, we can choose the basis in which they both are diagonal. In this basis

$$\|\rho_0^{\otimes N} - \rho_1^{\otimes N}\|_1 = \sum_{k=(k_1, \dots, k_d)} |p_1^{k_1} \dots p_d^{k_d} - q_1^{k_1} \dots q_d^{k_d}|, \quad (4.12)$$

where k is a partition of N , and (p_1, \dots, p_d) and (q_1, \dots, q_d) are eigenvalues of ρ_0 and ρ_1 , respectively. On the right-hand side we have $\|P_N - Q_N\|_1$, the distance between two multinomial distributions, P_N and Q_N , that arise in repeated trials from distributions P and Q . Therefore, because $R = \frac{1}{2} (1 - \frac{1}{2} \|\rho_0^{\otimes N} - \rho_1^{\otimes N}\|_1)$ in the quantum case and $R = \frac{1}{2} (1 - \frac{1}{2} \|P_N - Q_N\|_1)$ in the classical case, the average errors and their asymptotic growth rates are the same in the quantum and classical cases.

QED.

4.3 Bounds

Let us now derive some simple bounds on the error probability that follows from known inequalities. These bounds are useful because they are rather narrow and easy to compute. The first set of bounds follows from inequalities between quantum fidelity and probability of error. The second bound is only applicable to the asymptotic rate of error decline, and it follows from a quantum analog of Stein's lemma.

Recall that *fidelity* between two states is defined as follows:

$$F(\rho_0, \rho_1) = \text{tr} \sqrt{\sqrt{\rho_0} \rho_1 \sqrt{\rho_0}}, \quad (4.13)$$

where \sqrt{X} is the unique non-negative definite, Hermitian matrix Y such that $Y^2 = X$.

Theorem 4.4 *Probability of error for optimal test with joint measurement satisfies the following bounds:*

$$\frac{1}{2} \left(1 - \sqrt{1 - [F(\rho_0, \rho_1)]^{2N}} \right) \leq R \leq \frac{1}{2} [F(\rho_0, \rho_1)]^N. \quad (4.14)$$

Asymptotically,

$$2 \log F(\rho_0, \rho_1) \lesssim \frac{1}{N} \log R \lesssim \log F(\rho_0, \rho_1) \quad (4.15)$$

If ρ_0 is pure, $\rho_0 = |\psi_0\rangle \langle \psi_0|$, the probability of error satisfies a tighter upper bound:

$$R \leq \frac{1}{2} [F(\rho_0, \rho_1)]^{2N} = \frac{1}{2} \langle \psi_0 | \rho_1 | \psi_0 \rangle^N. \quad (4.16)$$

Proof: The first result follows from inequality (44) in Fuchs and van de Graaf (1999) applied to the case of the sample of N independent states, and from the fact

that $F(\rho_0^{\otimes N}, \rho_1^{\otimes N}) = [F(\rho_0, \rho_1)]^N$. The second one is a consequence of Exercise 9.21 in Nielsen and Chuang (2000). For the reader's convenience, I include below short proofs of these results.

The Fuchs-Graaf result states that for every pair of quantum states, ρ_0 and ρ_1 , it is true that

$$1 - F(\rho_0, \rho_1) \leq \frac{1}{2} \|\rho_0 - \rho_1\|_1 \leq \sqrt{1 - F(\rho_0, \rho_1)^2}. \quad (4.17)$$

These inequalities follow because

(1) $F(\rho_0, \rho_1) = \min_{P, Q} F(P, Q)$, where distributions P and Q arise from a measurement of states ρ_0 and ρ_1 , and where $F(P, Q) =: \sum_i \sqrt{p_i q_i}$;

(2) $\|\rho_0 - \rho_1\|_1 = \max_{P, Q} \|P - Q\|_1$, where P and Q come from a measurement, and where $\|P - Q\|_1 =: \sum_i |p_i - q_i|$;

(3) the corresponding inequality holds for probability distributions

$$1 - F(P, Q) \leq \frac{1}{2} \|P - Q\|_1 \leq \sqrt{1 - F(P, Q)^2}. \quad (4.18)$$

Indeed, given (1), (2), and (3), the left-hand inequality in (4.17) follows because

$$1 - F(\rho_0, \rho_1) \stackrel{(1)}{=} 1 - F(P, Q) \quad (\text{for certain } P \text{ and } Q) \quad (4.19)$$

$$\stackrel{(3)}{\leq} \frac{1}{2} \|P - Q\|_1 \stackrel{(2)}{\leq} \frac{1}{2} \|\rho_0 - \rho_1\|_1. \quad (4.20)$$

The right-hand inequality follows similarly.

Result (1) is from Fuchs and Caves (1995). (2) is a restatement of the Holevo-Helstrom result (4.1). The left-hand inequality in (3) holds because

$$\sum_i |p_i - q_i| \geq \sum_i (\sqrt{p_i} - \sqrt{q_i})^2 \geq 2 \left(1 - \sum_i \sqrt{p_i q_i} \right). \quad (4.21)$$

The right-hand inequality in (3) holds because

$$\left(\sum_i |p_i - q_i| \right)^2 = \left(\sum_i |\sqrt{p_i} - \sqrt{q_i}| |\sqrt{p_i} + \sqrt{q_i}| \right)^2 \quad (4.22)$$

$$\leq \sum_i |\sqrt{p_i} - \sqrt{q_i}|^2 \sum_i |\sqrt{p_i} + \sqrt{q_i}|^2 \quad (4.23)$$

$$= 4 \left(1 - \left(\sum_i \sqrt{p_i q_i} \right)^2 \right). \quad (4.24)$$

To prove the second part of the theorem, we need to prove that if $\rho_0 = |\psi_0\rangle \langle \psi_0|$, then there is such a measurement and a decision rule that

$$R \leq \frac{1}{2} \langle \psi_0 | \rho_1 | \psi_0 \rangle. \quad (4.25)$$

Take measurement $\{P_{\psi_0}, I - P_{\psi_0}\}$, where P_{ψ_0} is the projector on vector ψ_0 . Then the probabilities of the first and second outcomes are respectively 1 and 0 if the state is ψ_0 , and $\langle\psi_0|\rho_1|\psi_0\rangle$ and $1 - \langle\psi_0|\rho_1|\psi_0\rangle$ if the state is ρ_1 . Define the decision rule as follows: state ψ_0 is accepted if and only if the first outcome occurs. The expected error of this rule is $\frac{1}{2}\langle\psi_0|\rho_1|\psi_0\rangle$.

In the case of tensor powers (4.25) becomes

$$R \leq \frac{1}{2} \langle\psi_0|\rho_1|\psi_0\rangle^N. \quad (4.26)$$

QED.

Note that the lower bound of inequality (4.14) binds for pure states. This can be seen from (4.6) because for pure states $F(\rho_0, \rho_1) = |\langle\psi_0|\psi_1\rangle|$. The upper bound of inequality (4.14) binds for certain commuting operators.

Another bound follows from results by Ogawa and Nagaoka (2000). Define quantum relative entropy:

$$D(\rho_0||\rho_1) = \text{tr} [\rho_0(\log \rho_0 - \log \rho_1)]. \quad (4.27)$$

Then the following lower bound on the error rate holds.

Theorem 4.5 $\log \max \{D(\rho_0||\rho_1), D(\rho_1||\rho_0)\} \lesssim \frac{1}{N} \log R$

Proof: R is the average of error probabilities of two types. Say, $R = \frac{1}{2}R_1 + \frac{1}{2}R_2$. If both R_1 and R_2 satisfy the inequality, then R also does. Ogawa and Nagaoka, proved that if one of the error probabilities violates this inequality, then the other error probability must approach one as the sample size grows, so the inequality will hold for the average of the error probabilities, R .

QED.

For an example of two-dimensional states, the bounds are illustrated in Figures 1,2,3 and 4. The states in the example are linear combinations of the Pauli matrices:

$$\rho_0 = \frac{1}{2}(I + a\sigma_1), \quad (4.28)$$

$$\rho_1 = \frac{1}{2}(I + (b \cos \theta)\sigma_1 + (b \sin \theta)\sigma_2), \quad (4.29)$$

where

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}. \quad (4.30)$$

Figures 1 and 2 suggests that bound from Theorem 4.5 is a good estimate of the error if the sample size is small and underestimates the error if the sample size is large. Figures 3 and 4 suggest that bound from Theorem 4.5 is especially good when the hypotheses are close to completely mixed state, $\frac{1}{2}I$.

5 Separable Measurements

In the previous section we have seen that it is difficult to compute the optimal joint measurement because of the high dimensionality of the problem involved. Besides, even if the optimal joint measurement is found, it can have an enormous number of outcomes, so it is hard to realize it in the laboratory. In this section we turn our attention to separable independent measurements. The goal is to show that the efficiency of a separable measurement with a small number of outcomes is not much smaller than the efficiency of the optimal joint measurement.

Let us denote the probabilities of the i -th outcome as p_i and q_i depending on whether the state is ρ_0 or ρ_1 . The following theorem about optimal measurements holds:

Theorem 5.1 *All outcomes of an optimal measurement are projectors.*

Proof: Indeed, if the measurement includes an outcome, M_0 , that is not a projector then it can be represented as a sum of projectors with non-negative coefficients:

$$M_0 = \sum_{i=1}^n \alpha_i M_i. \quad (5.1)$$

Therefore

$$p_0 = \text{tr}(M_0 \rho_0) = \sum_{i=1}^n \alpha_i p_i, \quad (5.2)$$

$$q_0 = \text{tr}(M_0 \rho_1) = \sum_{i=1}^n \alpha_i q_i. \quad (5.3)$$

Since function $x^\lambda y^{1-\lambda}$ is concave and homogeneous, we have the following inequality

$$p_0^\lambda q_0^{1-\lambda} \geq \sum_{i=1}^n (\alpha_i p_i)^\lambda (\alpha_i q_i)^{1-\lambda}. \quad (5.4)$$

Because of (3.10), this inequality implies that we can decrease the error by using the set of outcomes $\{M_i\}$ instead of M_0 . This contradicts the optimality of the measurement.

QED.

How many outcomes does an optimal measurement have? It turns out that if one of the states is pure, $\rho_0 = |\psi_0\rangle\langle\psi_0|$, then only two outcomes is needed - a huge reduction relative to the d^N outcomes needed for the optimal joint measurement.

Theorem 5.2 *When one of the states is pure, there is an asymptotically optimal test with two outcomes in each measurement. The average error probability of the test satisfies the following bound*

$$R \lesssim \frac{1}{2} \langle\psi_0|\rho_1|\psi_0\rangle^N \text{ as } N \rightarrow \infty. \quad (5.5)$$

Proof: Take measurement $\{P_{\psi_0}, I - P_{\psi_0}\}$, where P_{ψ_0} is the projector on vector ψ_0 . Then the probabilities of the first and second outcomes are respectively 1 and 0 if the state is ρ_0 , and $\langle \psi_0 | \rho_1 | \psi_0 \rangle$ and $1 - \langle \psi_0 | \rho_1 | \psi_0 \rangle$ if the state is ρ_1 . Define the decision rule as follows: state ρ_0 is accepted if and only if the second outcome never occurred. This rule leads to an error if and only if the true state is ρ_1 and the second outcome never occurs. Thus the average probability of error for this decision rule is

$$R = \frac{1}{2} \langle \psi_0 | \rho_1 | \psi_0 \rangle^N. \quad (5.6)$$

Thus the rates of error decline coincide for the cases of joint and separable measurements. Since the optimal separable test cannot do better than the optimal joint measurement, the measurement considered is optimal.

QED.

If both states are mixed, then we can use the measurement that maximizes fidelity distance between distributions of outcomes. In other words, the measurement is chosen in such a way that it minimizes

$$F(P, Q) = \sum \sqrt{p_i q_i}. \quad (5.7)$$

We will call this measurement fidelity-optimal. The advantage of this method is that the fidelity-optimal measurement is easy to compute. It is simply a measurement with outcomes that are orthogonal projectors on the eigenvectors of the following operator:

$$M = \rho_1^{-1/2} \sqrt{\rho_1^{1/2} \rho_0 \rho_1^{1/2}} \rho_1^{-1/2}. \quad (5.8)$$

(See Fuchs and Caves (1995) for an explanation why this M is fidelity-optimal.) This measurement has only d outcomes and their probabilities are easy to compute. For this fidelity-optimal measurement we can write a bound on the asymptotic error:

Theorem 5.3 *The asymptotic error of the test based on the fidelity-optimal measurement has the following asymptotic bound:*

$$\frac{1}{N} \log R \lesssim \log F(\rho_0, \rho_1). \quad (5.9)$$

This is the same upper bound that we have for joint asymptotic measurement according to Theorem 4.4.

Proof:

$$\frac{1}{N} \ln R = \min_{0 \leq \lambda \leq 1} \log \sum_{i=1}^N p_i^\lambda q_i^{1-\lambda} \leq \log \sum_{i=1}^N \sqrt{p_i q_i} \leq \log F(\rho_0, \rho_1). \quad (5.10)$$

The equality holds because of (3.10), and the second inequality is inequality (44) in Fuchs and van de Graaf (1999).

QED.

6 Illustration

This section illustrates the concepts developed above with an example of testing for the presence of entanglement. Entanglement is one of the properties of quantum systems that clearly separates them from classical systems. It is a co-dependence of two remote parts of a quantum system that cannot be created or destroyed by local operations on the parts. Entanglement has become an important part of many quantum technologies including quantum teleportation and quantum cryptography.

Entanglement has been produced in the laboratory. For example, Turchette et al. (1998) developed a technique in which two ions are trapped and illuminated equally by a laser beam that results in the creation of entanglement.

In this illustration we are interested in tests of whether the entanglement has been produced or not.

An example of an entangled quantum state is a pure state of the system of two particles that corresponds to the projector on the following vector:

$$\psi_0 = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle), \quad (6.1)$$

where $|00\rangle$ and $|11\rangle$ denote $|0\rangle \otimes |0\rangle$ and $|1\rangle \otimes |1\rangle$, and $|0\rangle$ and $|1\rangle$ form an orthonormal basis in the Hilbert space corresponding to one of the particles.

The density matrix for this system is

$$\rho_0 = |\psi_0\rangle \langle \psi_0| = \begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}. \quad (6.2)$$

The alternative hypothesis is that the state is a mix of two non-entangled states given by projectors on vectors $|00\rangle$ and $|11\rangle$, respectively. The density matrix for this hypothesis is

$$\rho_1 = \begin{pmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \end{pmatrix}. \quad (6.3)$$

This state can be easily produced by local operations but it is useless for technologies that require entanglement.

Applying Theorem 5.2, we obtain the following formula for the asymptotic error

$$R \sim \frac{1}{2} \langle \psi_0 | \rho_1 | \psi_0 \rangle^N = \frac{1}{2^{N+1}}. \quad (6.4)$$

It follows that it is sufficient to measure a sample of size 3 to reduce error below 5%.

The components of the optimal separable measurement are the projection on ψ_0 and its complement. Note that this is a joint measurement of both particles. Actually, ρ_0 and ρ_1 cannot be distinguished by the measurements that operate

on each particle separately. This problem is statistically unidentified by local measurements.

7 Conclusion

We have estimated the Chernoff efficiency bound for cases of joint and separable measurements and also calculated it exactly for both pure and commuting states. The results suggest that the loss of efficiency caused by restriction to separable measurements is small.

Several questions remain open. Notably, it is not known whether the joint measurement can ever be asymptotically better than the optimal separable measurement. Second, it is not clear whether the optimal separable measurement consists of orthogonal projectors. Third, it is unclear whether the number of outcomes in this measurement is finite for finite-dimensional quantum states.

References

- Barndorff-Nielsen, O. E., R. D. Gill, and P. E. Jupp (2001). On quantum statistical inference. Working Paper, available at <http://www.math.uu.nl/people/gill/Preprints/qiread9statsoc.pdf>.
- Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics* 23(4), 493–507.
- Cirac, J. I. and P. Zoller (1995). Quantum computation with cold trapped ions. *Physical Review Letters* 74, 4091–4094.
- Cover, T. M. and J. A. Thomas (1991). *Elements of Information Theory*. John Wiley and Sons, Inc.
- Fuchs, C. A. and C. A. Caves (1995). Mathematical techniques for quantum communication theory. *Open Systems and Information Dynamics* 3(3), 345–356. also available at <http://arxiv.org/abs/quant-ph/9604001>.
- Fuchs, C. A. and J. van de Graaf (1999). Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory* 45(4), 1216–1227.
- Helstrom, C. W. (1976). *Quantum Detection and Estimation Theory*. Academic Press: New York.
- Hoeffding, W. (1965). Asymptotically optimal tests for multinomial distributions. *Annals of Mathematical Statistics* 36(2), 369–401.
- Holevo, A. S. (1976). *Investigation of a general theory of statistical decisions*. English translation: Proceeding of Steklov Institute of Mathematics, v.3, 1978.
- Holevo, A. S. (2001). *Statistical Structure of Quantum Theory* (1 ed.). Springer-Verlag. Lecture Notes in Physics. Monographs; 67.

- Nielsen, M. A. and I. L. Chuang (2000). *Quantum Computation and Quantum Information*. Cambridge University Press.
- Ogawa, T. and H. Nagaoka (2000). Strong converse and Stein’s lemma in quantum hypothesis testing. *IEEE Transactions on Information Theory* 46(3), 2428–2433.
- Parthasarathy, K. R. (2001). On consistency of the maximum likelihood method in testing multiple quantum hypotheses. In *Stochastics in finite and infinite dimensions*, Trends Math., pp. 361–377. Boston, MA: Birkhauser Boston.
- Sanov, I. N. (1957). On the probability of large deviations of random variables. *Matematicheskii Sbornik* 42, 11–44.
- Santaló, L. A. (1976). *Integral geometry and geometric probability*. Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam. With a foreword by Mark Kac, Encyclopedia of Mathematics and its Applications, Vol. 1.
- Turchette, Q. A., C. S. Wood, B. E. King, C. J. Myatt, D. Leibfried, W. M. Itano, C. Monroe, and D. J. Wineland (1998). Deterministic entanglement of two trapped ions. *Physical Review Letters* 81(17), 3631–3634.